# 商铺关联分析：

1. 使用Apriori原理得到商铺的最高M元频繁集
2. 使用Apriori原理得到商铺的每个频繁集中蕴含的关联规则
3. 基于关联规则得到商铺之间的有向依赖图（邻接表）
4. 将邻接表转化成邻接矩阵，然后转化为page rank转移矩阵
5. 根据page rank转移矩阵得到每个商铺的评分

涉及到的数据库表**mac\_portal：**

mac

portalId

## 使用Apriori原理得到商铺的最高M元频繁集

1. 生成商铺共现清单：

清单结构：

|  |  |
| --- | --- |
| {portalId列表} | 出现次数counter |

生成方法：

* 生成用户mac表mac：

Id// 自增id

mac

insert into mac(mac) select distinct mac from mac\_portal

* 并行生成清单结构，每个线程获取一部分mac，并执行：

select \* from mac\_portal where mac in {“xxxx”,”yyyy”....} group by mac, portalId

对相同的mac访问过的portalId组织成portalId列表，存储进清单结构

1. 进行频繁集搜索

初始化频繁集列表，每个频繁集只有一个portalId

I = 1;

While(i <= M) {

在清单中搜索每个频繁集出现的次数，并计算、保存其支持度，记为S

剔除频繁集列表中支持度小于阈值的频繁集

把频繁集列表中的频繁集组合出更大的频繁集，合并出的频繁集大小为i+1

++i

}

## 使用Apriori原理获得每个频繁集中的关联规则

关联规则是频繁集的一个划分

对于每一个频繁集{p1, p2, p3}

对于p1,p2->p3的可信度，计算方式是S(p1p2p3)/S(p1p2)

假如p1,p2->p3的可信度小于阈值，则不用考虑对于频繁集划分左部是p1p2子集的情况